Global Age Mapping Integration (GAMI)
(https://doi.org/10.5880/GFZ.1.4.2023.006)

Simon Besnard[1*], Maurizio Santoro[2], Martin Herold[1], Oliver Cartus[2], Jonas Gütter, Viola H. A. Heinrich[1], Bruno Herault[3], Justin Kassi[4], Sujan Koirala[5], Anny N'Guessan[4], Christopher Neigh[6], Jacob A. Nelson[5], Benjamin Poulter[6], Ulrich Weber[5], Tao Zhang[7], and Nuno Carvalhais[5*]

1. *GFZ German Research Centre of Geosciences, Potsdam, Germany*
2. *Research Unit Forests and Societies, CIRAD, University of Montpellier, Montpellier, France*
3. *Laboratoire de Botanique, UFR Biosciences, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire*
4. *Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany*
5. *Laboratoire de Botanique, UFR Biosciences, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire*
6. *Biospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA*
7. *Department of Biology, University of Florida, Gainesville, Florida, USA*

***Contact persons:** Simon Besnard (besnard@gfz-potsdam.de) and Nuno Carvalhais (ncarvalhais@bgc-jena.mpg.de)

1. Licence

2. Citation

When using the data please cite:

Besnard, S.; Santoro, M.; Herold, M.; Cartus, O.; Gütter, J.; Heinrich, V. H.A.; Herault, B.; Kassi, J.; Koirala, S.; N'Guessan, A.; Neigh, C.; Nelson, J. A.; Poulter, B.; Weber, U.; Zhang, T.; Carvalhais, N. (2024): Global Age Mapping Integration (GAMI). V. 2.0. GFZ Data Services. https://doi.org/10.5880/GFZ.1.4.2023.006

The data are supplementary material to:

Besnard, S.; Koirala, S.; Santoro, M.; Weber, U.; Nelson, J.; Gütter, J.; Herault, B.; Kassi, J.; N'Guessan, A.; Neigh, C.; Poulter, B.; Zhang, T.; & Carvalhais, N. (2021). Mapping global forest age from forest inventories, biomass and climate data. In Earth System Science Data (Vol. 13, Issue 10, pp. 4881–4896). Copernicus GmbH. https://doi.org/10.5194/essd-13-4881-2021

Table of contents

3. Data Description

This dataset is an updated version of the MPI-BGC forest age product (Besnard, 2021 (https://doi.org/10.17871/ForestAgeBGI.2021), providing global distributions of forest age for 2010 and 2020 at a pixel size of 100 meters. The updated forest age product is now named the Global Age Mapping Integration (GAMI). Our approach leverages machine learning, specifically XGBoost, to generate data-driven estimates of forest age. These estimates are derived from a comprehensive dataset encompassing over 40,000 forest inventory plots, biomass/height measurements, remote sensing observations, and climate data.

One substantial update in our product is combining Landsat-based disturbance history with machine learning-based forest age estimates.

The age maps were created using a model that incorporates the 20 realisations of the biomass data, addressing aleatoric uncertainties. Additionally, we trained multiple XGBoost models with different optimal hyperparameters to capture the variability in model performance and predictions, thus accounting for epistemic uncertainties. By using these different biomass maps and a range of hyperparameter settings, we were able to quantify both types of uncertainties in the forest age estimates.

We generated 20 realisations of the AGB data by introducing controlled perturbations to the mean biomass estimates. Specifically, for each time step and each member of our ensemble, we added a scaled version of the standard deviation of the biomass, with the scaling factor drawn from a normal distribution clipped to the range [-1, 1]. This approach allows us to simulate the natural variability and uncertainty in the biomass data. We then combined these perturbed maps to create a comprehensive dataset representing the possible range of biomass values. This enables us to provide confidence intervals and robust statistical measures of uncertainty in our estimates. The following equation can describe the perturbation applied:

$$AGBi,j,k(t) = max(i,j(t) + Ski,j(t), 0) \qquad \text{(eq. 1)}$$

where AGBi,j,k(t) is the perturbed aboveground biomass at location (i,j) and time t for the kth member, i,j(t) is the mean aboveground biomass, i,j(t) is the standard deviation of the aboveground biomass, and Skis the scaling factor drawn from a normal distribution N(0,13) and clipped to the range [−1,1]. The use N(0,13) ensures that the perturbations are centred around zero, introducing symmetric and controlled variability. The standard deviation 13 moderates the variability to prevent extreme perturbations, thus maintaining the realism of the biomass estimates. Clipping the values to [−1,1] further ensures that the introduced variations are within a reasonable range.

Additionally, we provide two age class fraction products at 0.25 and one-degree spatial resolution using two-decade intervals up to 200 years, then 201-300, and >300 age classes. Custom spatial resolutions and age class specifications are available upon request.

4. Methodology

4.1. Mapping

The methodology for mapping the time since the last disturbance and afforestation using Landsat-based products is based on the studies by Potapov et al. (2021) and Hansen et al. (2013). The methodology involves several steps:

Create Last Time Since Disturbance Layer: This step begins by identifying disturbed forest areas using the layer "forests affected by stand-replacement disturbances or degradation" and initializing an image for the last time since disturbance. The Hansen data (https://storage.googleapis.com/earthenginepartners-hansen/GFC-2022-v1.10/download.html) data determines forest loss years, subtracted from 20 to obtain the time since the last disturbance in 2020. Negative values are masked as they refer to the losses in 2021 and 2022.

Create Stable Forest Layer: In this step, stable forest areas are identified using the layer 'Stable forest extent' from Potapov data and assigned a value of 21.

Create Gain Forest Layer (Afforestation): Afforested (regrown) forest areas are identified using the Forest extent gain layer and the canopy height 2000 from Potapov data. The afforested forest areas are divided into two categories based on forest height, with a value of 19 if the canopy height 2000 <5m and 20 if the canopy height 2000 >=5m.

Merge All Forest Age Layers: The previous layers representing stable, disturbed, and afforested forests are merged into a single layer by adding them, with masking applied to maintain only valid forest age values. The resulting layer represents Landsat-based last time since disturbance and afforestation information.

Resample and Reproject: Finally, the combined layer is resampled to a 100m pixel size using mode resampling and reprojected to the EPSG:4326 projection. This ensures consistency in the spatial resolution and projection of the machine learning-based forest age product. If most 30m pixels with a 100m pixel are non-forest (i.e., no-data value), then the 100m pixel is classified as non-forest.

4.2. Landsat-based time fusion since the last disturbance and afforestation products

The fusion of the Landsat-based time since the last disturbance and afforestation products with the machine learning-based forest age product is done following three steps:

1. Suppose Landsat-based estimates are less than or equal to 19 and the machine-learning predicted forest age is higher than Landsat-based estimates. In that case, we assign the forest age from Landsat-based estimates.
2. Suppose Landsat-based estimates are less than or equal to 19 and the machine-learning predicted forest age is lower or equal than Landsat-based estimates. In that case, we assign the forest age from the machine-learning estimates.
3. Suppose Landsat-based estimates are more than or equal to 20. In that case, we assign the forest age from the machine-learning estimates regardless of whether the machine-learning estimates are lower or higher than 20.

This fusion improves the forest age product by addressing biases of overestimating young forests that have experienced stand-replacement followed by regrowth or afforestation. However, some residual biases may still exist.

5. File description

- The data are provided in NetCDF format
- Spatial Reference System: EPSG 4326

Interactive Exploration: Explore the the 100m forest age map online may be (https://besnardsim.users.earthengine.app/view/globalforestage)

5.1. File inventory

Forest age product at 100m pixel

- GAMIv2-0_2010-2020_100m.nc

age class fraction products

- GAMIv2-0_2010-2020_class_fraction_0deg083.nc
- GAMIv2-0_2010-2020_class_fraction_0deg1.nc
- GAMIv2-0_2010-2020_class_fraction_0deg25.nc
- GAMIv2-0_2010-2020_class_fraction_0deg5.nc
- GAMIv2-0_2010-2020_class_fraction_1deg.nc

5.2. Description of variables

GAMIv2-0_2010-2020_100m.nc

Variable name	**unit**	**Description**
forest_age	years	Forest age using a fusion machine learning and Landsat-based last time since disturbance

This dataset contains forest age data in years at 100m pixel for the year 2010 and 2020.

Dimensions and Coordinates

- **Dimensions**: The dataset has four dimensions:
    - **Members**: 20 members
    - **Latitude**: ranging from 89.5 to -89.5 degrees
    - **Longitude**: ranging from -179.5 to 179.5 degrees
    - **Time**: Two time points (2010-01-01 and 2020-01-01)

Data Variables

- **forest_age**: (members, latitude, longitude, time) - Contains forest age data in years.

File Format

The dataset is stored in NetCDF format, ensuring compatibility with a wide range of geospatial and scientific data processing tools.

GAMIv2-0_2010-2020_class_fraction_*.nc

Variable name	**unit**	**Description**
forest_age	adimensonial	Fraction of each forest age class

These data represent the fraction of each age class using two-decade intervals up to 200 years, then 201-300, and >300 age classes at a series of pixel size: 0.083 degrees, 0.1 degrees 0.25 degrees, 0.5 degrees and one-degree pixel sizes.

Dimensions and Coordinates

- **Dimensions**: The dataset has five dimensions:
    - **Members**: 20 members
    - **Age Class**: Describes the age ranges of the forests in two-decade intervals up to 200 years, then 201-300, and >300 years.
    - **Latitude**: ranging from 89.5 to -89.5 degrees
    - **Longitude**: ranging from -179.5 to 179.5 degrees
    - **Time**: Two time points (2010-01-01 and 2020-01-01)

Data Variables

- **forest_age**: (members, age_class, latitude, longitude, time) - Contains the fraction of each age class. The sum of all the fraction is equal to one.

File Format

The dataset is stored in NetCDF format, ensuring compatibility with a wide range of geospatial and scientific data processing tools.

6. Limitations and recommendations

The integration of Landsat-based data with machine learning (ML) estimates enhances the accuracy of determining the time since the last stand-replacement event over the past 20 years. This approach mitigates certain biases associated with overestimating the age of young forests that have undergone stand replacement followed by regrowth or afforestation. However, the dataset may still overestimate the age of young forests and underestimate the age of relatively older forests. To limt those biases in forest age estimates, we recommend binning the data into intervals of at least two decades (as done in the GAMIv2-0_2010-2020_class_fraction_*.nc files). This approach accounts for potential inaccuracies in absolute forest age estimates at a 100-meter resolution. Additionally, a validation exercise will be necessary once a comprehensive set of national forest inventory data with precise coordinates is collected. This validation will further ensure the reliability and accuracy of the dataset.

7. References

Besnard, S., Koirala, S., Santoro, M., Weber, U., Nelson, J., Gütter, J., Herault, B., Kassi, J., N'Guessan, A., Neigh, C., Poulter, B., Zhang, T., and Carvalhais, N. (2021). Mapping global forest age from forest inventories, biomass and climate data. Earth System Science Data. https://doi.org/10.5194/essd-13-4881-2021

Besnard, S. (2021). The MPI-BGC 1km global forest age datasets (Version 1.0.0) [Data set]. Max Planck Institute for Biogeochemistry, Jena. https://doi.org/10.17871/FORESTAGEBGI.2021

Hansen, M. C. et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 850-853 (2013). https://doi.org/10.1126/science.1244693

Potapov, P. et al. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sens. Environ. 253, 112165 (2021). https://doi.org/10.1016/j.rse.2020.112165